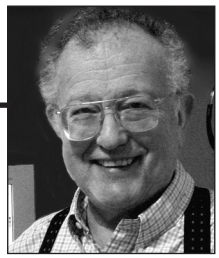
CPA Donald O. Hebb Award For Distinguished Contributions to Psychology as a Science – 2004
Prix Donald O. Hebb pour contributions remarquables à la psychologie en tant que science – 2004

Auditory Scene Analysis and the Role of Phenomenology in Experimental Psychology

ALBERT S. BREGMAN
McGill University



Abstract
This paper is about two topics: a) the role of subjectivity in psychological research; and, b) my research on the perceptual organization of sound, in which subjectivity has played an important role. Audio demonstrations that appeal to the subjective experience of the reader are presented in lieu of objective research data to support the claims made about auditory organization. It is argued that all psychological research depends on an underlying framework of intuition (unformalized knowledge acquired through everyday experience), and that this intuition plays a role in the design of the experiment.

Personal Experience as Data

The personal experience of the researcher has not fared well in scientific psychology. Since the failure of Titchener's Introspectionism in the early 20th century, and the rise of Behaviourism, scientific psychology has harboured a deep suspicion of the experience of the researcher as an acceptable tool in research. One would think that the study of *perception* would be exempt from this suspicion, since the subject matter of the psychology of perception is supposed to be about how a person's experience is derived from sensory input. Instead, academic psychology, in its behaviouristic zeal, redefined perception as the ability to respond differently to different stimuli – bringing it into the stimulus-response framework. Despite Behaviourism's fall from grace, psychology still insists on a behaviouristic research methodology.

In my own research, however, subjectivity has played a central role. It was a perceptual experience that got me going on the topic of perceptual organization in the first place: I was preparing an experiment on learning, involving a rapid sequence of unrelated sounds, each about the length of a speech phoneme. I spliced together one-tenth-second segments of many different sounds – water splashing in a sink, a dentist's drill, a tone, a vowel, etc. When I played the tape back to myself, though, I did not experience the sequences in the order that they were recorded on the tape. It appeared that nonadjacent sounds were grouping together and appeared to be adjacent. It was the similar sounds that seemed to be forming integrated perceptual sequences. This reminded me of an essay I had written at the University of Toronto on the topic of Gestalt Psychology. Some of the Gestaltist examples showed that similar visual forms would group together and segregate from dissimilar ones. Perhaps an analogous sort of grouping might be happening in my auditory sequence. Although I had never been trained in auditory perception research, this one subjective experience set me off on a 36-year period of study.

When I use the term "phenomenology," it is not with the technical meaning it has in the writings of Husserl or Heidegger, but is just a fancy name for experience. In all my years of research on the conditions under which a mixture of sounds will blend or be heard as separate sounds, my own phenomenology has played a central role in deciding what to study and how to study it. Also, the subjective experiences of colleagues and students have made it possible for them to understand the phenomena by listening to auditory demonstrations. I almost never carried out a study whose outcome I did not know in advance by listening to the stimuli. Only when I had figured out the conditions that would give rise to the effect I wanted to study, would I design a formal experiment.

It is impossible to overestimate how many years of research this saved. It made it possible to study a large set of theoretical issues in perceptual organization without elaborate sets of preliminary experiments to establish the right parameters. The approach of listening to many variants of the signals and getting familiar with their effects at a personal level permitted us to speed up the development of a general overview of auditory grouping (Bregman, 1990/1994), rather than to merely accumulate more and more highly quantitative knowledge about a narrow experimental paradigm – a not-unknown practice in experimental psychology. In the language of artificial intelligence, our approach would be called "breadth-first search" as contrasted with "depth-first search."

Canadian Psychology/Psychologie canadienne, 2005, 46:1, 32-40

Yet the role of subjectivity has often been criticized by journal reviewers. In the reviews of my first published article on auditory stream segregation, which showed that a rapid alternation of high and low sounds segregated into two perceptual streams, one of the sceptical reviewers proposed that there was something wrong with my loudspeakers – perhaps they continued to give out sound after the tone went off – and insisted that I test them. I was convinced that if the reviewers had merely listened to the sounds, their objections would have evaporated, but in those days you did not send in audio examples with your manuscript. I am not sure it would be acceptable for most journal editors even today. That is where the study of vision has an edge – it has always been possible to include visual illustrations. Since I cannot include audio demonstration with this article, I will refer to the demonstrations on a compact disk put out by our laboratory (Bregman & Ahad, 1996). I will refer to it, using abbreviations, for example, "B&A #1" to represent "Bregman and Ahad (1996), Demonstration Number 1."

The demonstration I would have included for the journal editors of my first article is B&A #1. It consists of a repeating cycle of six tones, three different high ones (H1, H2, H3) and three different low ones (L1, L2, L3), in the order H1, L1, H2, L2, H3, L3 (repeated many times). At a slow speed the tones are heard in the order in which they occur, but at high speed (say 100 ms per tone, onset to onset time) the listener hears two distinct streams of sound, one formed by the high tones and a second formed by the low ones (i.e.,

H1 - H2 - H3 - H1 - H2 - H3 - … etc.,
and L1 - L2 - L3 - L1 - L2 - L3 -,… etc.)

I got around the editorial taboos concerning subjective experience by giving many talks accompanied by taped auditory examples and eventually by publishing my own compact disk of auditory demonstrations (Bregman & Ahad, 1996). However, the CD did not come until 23 years after the first research paper. Nowadays one can put demonstrations on the web right away and refer reviewers to the website. The only problem is that websites are not archival in nature. They can easily disappear. Thus, there is a need for the publication of compact disks.

Something else that reviewers have criticized is the use of a subjective rating scale, in which listeners are asked, for example, to rate how clearly they can hear a sound in a mixture. However, this method leads to results that are reliable (statistically strong) and that can be predicted from theoretical considerations – the ultimate test of a measure's validity. Psychology journals on the whole prefer tasks that involve accuracy: You should be able to score the answers of the subjects as either correct or incorrect (e.g., by asking whether a particular sound was or was not present in a mixture of sounds), rather than simply accepting the participants' answers when they rate the clarity with which a target sound was heard. Sometimes we have used both types of measures either in the same experiment or in a pair of related experiments. The two types have given similar results, but the subjective rating scales have been more sensitive.

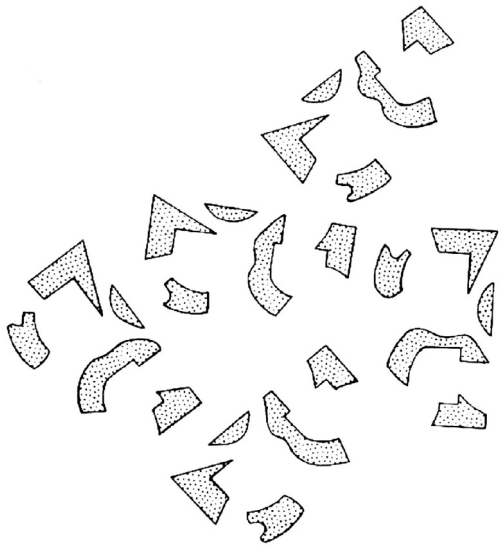


Figure 1. Fragments of a visual display (Bregman, 1990/1994, Figure 1.13). Reprinted with permission from MIT Press.

The Scene Analysis Problem

In this paper, I describe my research on auditory perception, but I do not present any data. Instead, I am going to support my arguments with audio demonstrations. Demonstrations are an appeal to the subjective experience of the listener. One of the main legacies of the Gestalt psychologists was their convincing visual demonstrations. After you have seen one, the question is "What causes it?" not "How many subjects were used, and what is the *p* value?" The auditory examples that I will use are also convincing, and give clear effects in a quiet room, or over headphones. (These are available on my website *www.psych.mcgill.ca/labs/auditory/laboratory.html* by clicking on "*Compact disk of demonstrations of auditory scene analysis.*" The desired demonstration can then be selected from the *List of Demonstrations.* Click on any demonstration that says "*Demo sample available*" and you will have the opportunity to play it. A visual illustration and an explanation of the demonstration can also be found

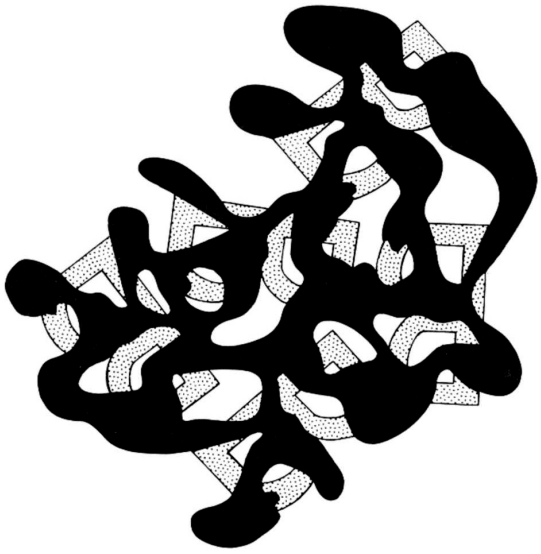


Figure 2. Same as Figure 1, but with the missing inkblot superimposed. A number of Bs now become visible (Bregman, 1990/1994, Figure 1.14). Reprinted with permission from MIT Press.

on the demonstration's webpage.) There are other demonstrations available, not just the ones for the present paper. For a larger set, see the CD by Bregman & Ahad (1996). The research to which these demonstrations are related is described in Bregman (1990/1994), which reviews all of the research on auditory scene analysis up to 1987. The booklet with the CD by Bregman and Ahad contains full references for the demonstrations that it presents.

Although this article is about auditory perception, I start with a visual demonstration. Figure 1 shows a set of fragments that are not interpretable as anything familiar. They were created by drawing a picture and then laying an amorphous inkblot on top of it as an "occluder." Then the parts covered by the inkblot were trimmed away, leaving the fragments shown in the figure, and the inkblot was removed. We can restore the perception of the underlying picture by simply putting the inkblot back on top, as in Figure 2. This does not supply any of the contours that were trimmed away, but it still makes it possible to see what the original picture was. The inkblot is simply seen as hiding some of the underlying picture, but revealing enough of it to see what it is. This is an example of Gestalt completion.

I introduced this example because a similar organizational principle is found in sound (Dannenbring, 1976); an example of it is illustrated visually in Figure 3 and in audio on B&A #29, Part 1. We start with a tone that glides up and down in frequency repeatedly. Then we remove a bit from each rising and each falling portion and replace it with a silent gap, causing discontinuities to be *heard* in the gliding tone, and *seen* in Figure 3, Panel 1. However, when loud noise bursts are inserted where the gaps were, as in Panel 2, the tone is heard as complete, gliding right through them. As in the visual example, the presence of an "occluder" – in this case a sound that might have obliterated parts of the tone – is interpreted as hiding the tone, and the brain restores what it predicts to be missing.

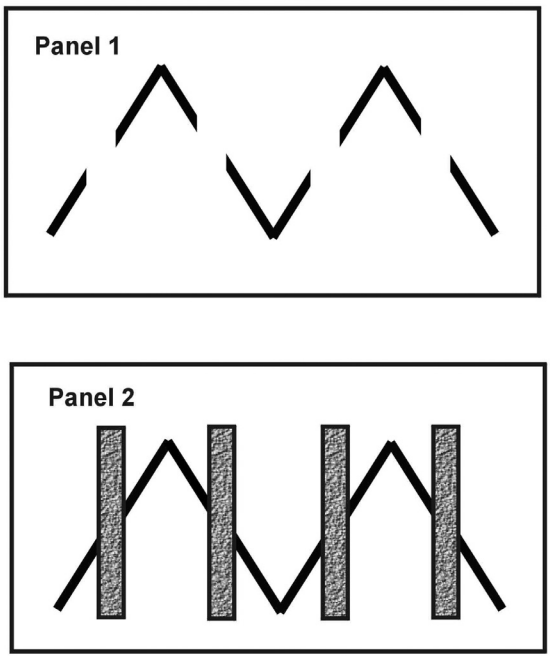


Figure 3. A tone gliding upward and downward in frequency is interrupted by silences in Panel 1. It is heard as discontinuous. When loud bursts of noise are inserted in the gaps, as in Panel 2, the tone is heard as continuously gliding through the noise bursts (adapted from Dannenbring, 1978, Figure 1). Reprinted with permission from CPA.

How can we account for the similarities of these two examples (i.e., the visual and the auditory example)? Is it just that the same Gestalt principles of grouping exist across all the senses? This may or may not be true, but there is a deeper question to be asked: Why should Gestalt principles of organization exist at all in perception? Answering this question requires a detour into the realm of machine intelligence.

Visual Scene Analysis

During the 1960s, researchers at Massachusetts Institute of Technology attempted to get computers to recognize visual forms, and started by projecting a picture on an array of sensors (Winston, 1975). It soon became apparent that the task needed to be simplified; so they used simple line drawings of a pile of reg-

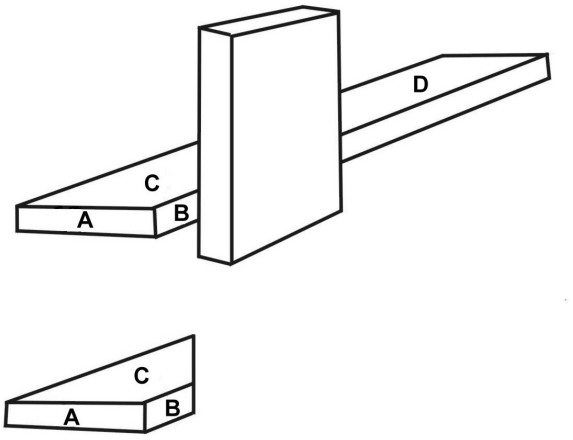


Figure 4. A line drawing of blocks for visual scene analysis (a simplification of the drawings used by Guzman, 1969).

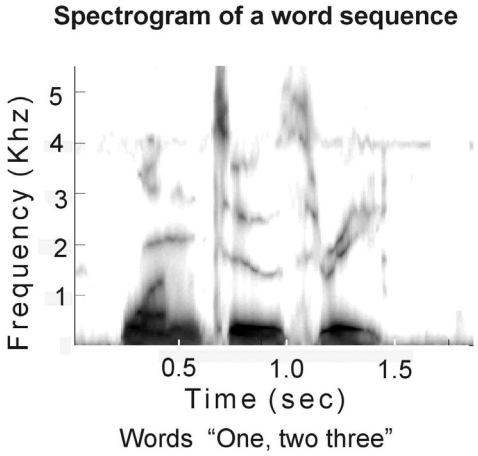


Figure 5. Spectrogram of a word sequence, "one, two, three."

ularly shaped blocks (Guzman, 1969). Figure 4 shows a simplified version of such a drawing. The task of the computer was to describe the shape of each block.

This should be easy. After all, according to the rules of drawing, the lines represent the edges of blocks. If an area is totally enclosed by lines, it is either a surface of a block or a hole formed by a set of surrounding blocks. However, there is a difficulty for the computer that is not apparent to a human viewer. Consider the adjacent areas labelled A, B, and C. We know by the rules of drawing that they are separate surfaces, but do they constitute a complete object? The human eye says no (D must also be included), but how could a computer tell that surfaces C and D were parts of the same object? The human eye says that it is so, but a computer would need some special rules to come up with that answer.

These problems led to the conclusion that a computer process would have to be designed that would do the equivalent of taking a crayon and colouring in, with the same colour, all surfaces of the same object. (Remember this "crayon." It is important in audition as well.) This would allow the shape-recognition process to focus only one object at a time. This process, called "scene analysis," was critical for achieving correct descriptions of the objects. If done incorrectly, the resulting descriptions might not correspond to the actual objects in the picture (e.g., in Figure 4, if only the regions A, B, and C were labelled as parts of an object, a recognizer would see the shape shown below the blocks).

But how does this relate to sound? I will take a detour to illustrate that there is a corresponding scene analysis problem in audition. Imagine a game played at the side of a lake. Two small channels are dug, side by side, leading away from the lake, and the lake water is allowed to fill them up. Part way up each channel, a cork floats, moving up and down with the waves. You stand with your back to the lake and are allowed to look only at the two floating corks. Then you are asked questions about what is happening on the lake. Are there two motorboats on the lake or only one? Is the nearer one going from left to right or right to left? Is the wind blowing? Did something heavy fall into the water? You must answer these questions just by looking at the two corks. This would seem to be an impossible task. Yet consider an exactly analogous problem. As you sit in a room, a lake of air surrounds you. Running off this lake, into your head, are two small channels – your ear canals. At the end of each is a membrane (the ear drum) that acts like the floating corks in the channels running off the lake, moving in and out with the sound waves that hit it. Just as the game at the lakeside offered no information about the happenings on the lake except for the movements of the corks, the sound-producing events in the room can be known by your brain only through the vibrations of your two eardrums. But the sense of hearing finds it easy to answer the same kinds of questions asked at the lakeside: Are there two talkers in the room or only one? Is the nearer one moving from left to right or right to left? Is the kettle hissing? Did something heavy just fall on the floor? These "easy" questions are exactly comparable to the ones asked at the lake. How can the ears answer questions that the eyes cannot? The difficulty is similar in the two cases. The movements of both the corks and the eardrums are determined by the sum of all the waves that enter the channels. The sum is just another wave pattern that does not have, written anywhere on it, that it is actually a sum of a set of waves, or what the component waves might be.

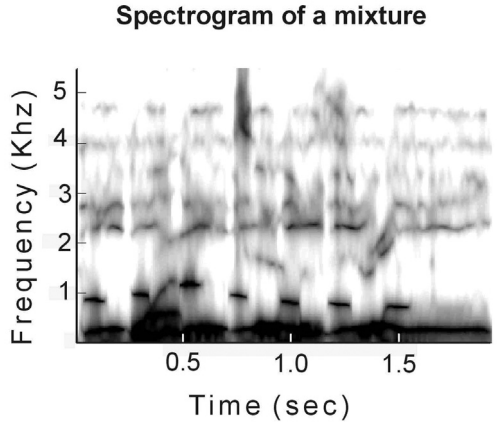


Figure 6. Spectrogram of a mixture: (a) the words "one, two, three," (b) singing "da-da-da," (c) whistling, (d) computer fan (Bregman & Woszczyk, 2004, Figure 1). Reprinted with permission from A. K. Peters.

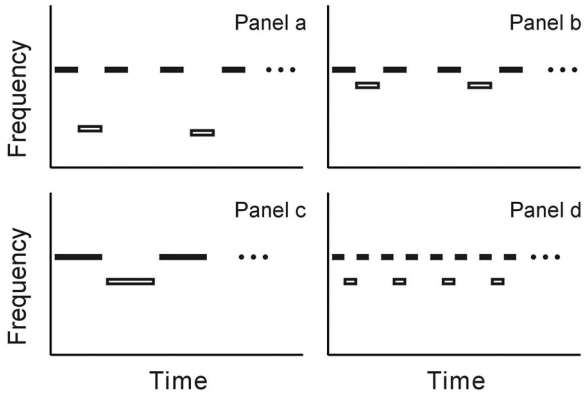


Figure 7. Repetitions of a high (H) and a low (L) tone in a galloping rhythm. Panel a: H and L are far apart in frequency. Panel b: H and L are close together in frequency. Panel c: Slow rate. Panel d: Fast rate (Bregman & Woszczyk, 2004, Figure 2). Reprinted with permission from A. K. Peters.

The auditory example can be better understood by converting the waveform of the sound into a *spectrogram*. The one in Figure 5 portrays the word sequence, "one, two, three." A spectrogram resembles a musical score: The vertical axis represents frequency and the horizontal axis represents time. However, unlike a musical score, in which only the fundamental frequency (pitch) of each sound appears on the vertical axis, the spectrogram shows all the frequency components of every sound. Most natural sounds contain many frequency components.

In a spectrogram of a real-life mixture of sounds, frequency components from different sounds are overlaid in top of one another. To get a spectrogram of a mixture (e.g., Figure 6), we would have to draw the spectrograms of each of the sounds of the mixture on separate pieces of transparent plastic, and then stack them on top of one another and view the result, in which the individual spectrograms were no longer evident. To make the latter more visible, it would be desirable to use a crayon to colour in, with the same colour, all those frequencies in the mixture that have been provided by the same sound (the same crayon used in the example of visual scene analysis). If this were done incorrectly, an acoustic component of one real-world sound might be heard as a part of another sound.

Horizontal (Sequential) and Vertical (Spectral) Dimensions of Organization

When we examine the spectrogram's representation of a mixture of sounds, the problem of perceptually separating the various component sounds can be seen to have two dimensions, the vertical and the horizontal. The problem associated with the vertical dimension involves grouping together the particular set of frequency components that came from the same environmental sound, from among all the ones present at the same time, and to do the same for each of the concurrent environmental sounds. The problem associated with the horizontal dimension involves grouping those frequency components that have come from the same environmental sound over time; we can call this "sequential grouping."

Sequential Grouping

We can study sequential organization in a simplified form by using a sequence of pure tones (see Figure 7 and listen to B&A #3). One demonstration consists of a repeated alternation of high (H) and low (L) tones in a galloping rhythm: HLH–HLH–HLH… (where the dash represents a silence equal in duration to a tone). We start with a sequence in which the H and L tones are far apart in frequency (Panel a) and gradually speed it up. At slow speeds we hear the galloping rhythm formed of the H and L tones, but as the sequence speeds up, we perceive two separate streams of sound, a higher stream containing only the H tones, and a lower one containing only the L tones. The galloping rhythm disappears and is replaced by separate regular repetitions of a single tone, the L in the low stream, and the H in the high stream. However, if the H and L tones are close together in frequency (Panel b), then even at high speeds, the galloping rhythm is still perceived and there is no segregation into high and low streams.

One can interpret the effects of speed as bringing each H tone closer to the next H tone, and each L

tone closer to the next L tone. Compare Panel c (slow) with Panel d (fast), with the same frequency separation. Think of each panel as a two-dimensional surface on which the tones are laid out. Both time and frequency contribute to the "distance" between pairs of tones. Tones that are closer to one another on this surface tend to group together. At low speeds (Panel c), each H tone is closer to the following L tone than it is to the following H tone; so it groups with the L tone. As the sequence speeds up (Panel d), each H tone comes closer in time to the next H tone and groups with it, so that the net frequency-by-time distance favours its grouping with the next H in preference to the next L. The eye, looking at Panel d, sees the same two groupings of the horizontal bars that represent the tones. We could just as easily have brought the sounds closer together in time by keeping their tempo constant but increasing the length (duration) of each tone.

This is an example of sequential grouping, since no two sounds are present at the same time. It shows that there is a tendency for similar sounds to group together to form streams and that both nearness in frequency and in time are grounds for treating sounds as similar. The Gestalt psychologists had shown that, in vision, objects that are nearer in space (or more similar) tend to form tighter perceptual clusters. The same principle seems to apply to audition.

The preceding example used the pitch of pure tones (based on their frequencies) as the variable that defined similarity, but there are many other ways in which short simple sounds can be similar or dissimilar. Among them are: 1) timbre (differences in the sound quality of tones despite identical pitches and loudnesses) – note that the difference between the vowel sounds "ee" and "ah" can be thought of as a timbre difference; 2) spectral similarity (i.e., to what extent they share frequency components [e.g., for noise bursts that have no pitch]); 3) temporal properties, such as the abruptness of onset of sounds; 4) location in space; and, 5) intensity.

One can think of each sound as seeking ones with which it is *most similar* and as forming clusters, or streams, based on this similarity. This means that a particular sound, A, may group with another sound, B, in one context, where there are no "better" sounds for each of them to group with, but each of these two sounds may group with another sound, X or Y, in a context in which X is very similar to A and Y is very similar to B (demonstrated in B&A#15). The competition is based on an overall similarity in which the properties that define similarity may be any combination of the ones that I listed in the previous paragraph, different properties having different weights in defining the overall similarity. This process of grouping can be seen as conforming to the Gestalt psychologists' observation that perception tends to create the "best" or "simplest" figures that it can.

When the galloping sequence breaks apart into two perceived sequences, a high one and a low one, we say that a single *auditory stream* has split apart into two *streams*. In vision, we refer to the result of grouping as an *object* (when the result is perceived as a unit), or as a *perceived group of objects* (when the result is a cluster of separate objects). In hearing, we refer to the result of auditory grouping as an *auditory object* or a *perceived sound* (when it creates a single sound), and as an *auditory stream* (when it creates a sequence that unfolds over time). The perception of a stream is the brain's way of concluding (correctly or incorrectly) that sounds included in the stream have been emitted over time by the same sound source (e.g., a drum, a voice, or an automobile).

Effects of Sequential Integration

Even more interesting than the factors that lead to segregation of sounds into separate auditory streams are the effects that this has on our perceptual experience of the sound. The first of these effects is that *fine details of temporal order* are available to our perception only when they concern sounds in the same stream. For example, when the galloping sequence, HLH–HLH–..., segregates very strongly into two streams, it is difficult to judge whether the L tone occurs exactly halfway in time between the two H tones. A second factor that emerges from the organization into streams is *rhythm* (which tends to be formed by sounds in the same auditory stream). In the galloping pattern, described earlier, the triplet rhythm of the H_L_H gallop was audible only so long as H and L were perceived to be in a single stream. A third result of the formation of streams is *melody*. Melodies tend to emerge from tones perceived as being in the same auditory stream. For example, when the galloping sequence split into two streams, the up-and-down (simple) melody of the gallop was lost and replaced by two streams, each of which contained only a simple pitch. This means that when a composer wants the listener to hear more than one melody (each in a different pitch range), the pitch ranges must be well separated. If they draw close together, a note may be "captured" perceptually into the wrong melody. You can listen to an illustration of how a melody depends on all its notes being in the same stream, isolated from other sounds that might also be present (B&A#5). Suppose we take a simple melody, "Mary Had a Little Lamb," and insert a note of a randomly chosen pitch between every two notes of the

melody (e.g., M 9 A 8 R 2 Y 3 H 8 A 6 D 4 A 2 L 8 I 7 T 4 … . [where the letters represent the notes of the melody and the digits represent randomly chosen distractor notes]). If the distractor notes are chosen from the same pitch range as the melody's notes, the melody is impossible to hear. However, if the distractor notes fall outside the range of the melody (say an octave above it) and form a separate stream, the melody can be heard clearly in its own stream.

Simultaneous Organization

I have, thus far, described the grouping of sounds that occur in a sequence, but the auditory system must also deal with environmental sounds that overlap in time. When the signals travelling from ear to brain represent a mixture of sounds that are present at the same time, the auditory system must sort out this information into a set of concurrent streams. If we re-examine the spectrogram of a mixture shown in Figure 6, we see that a vertical slice contains more than one frequency (a single frequency would be shown as a single thin horizontal line). Yet it is not immediately obvious how the frequencies in this slice should be allocated as components of various concurrent sounds.

There are many cues in the sound mixture that help the auditory system group the frequency components that have arisen from the same environmental sound. Space limitations prevent me from mentioning all of them here but the cue of *synchronized changes* represent an example that cuts across sensory modalities. In vision, suppose I am out on a sunny day and see a flock of flying birds crossing the path of another flock moving in a different direction. Although the birds may be too far away to be seen as distinct shapes, my visual system tends to group the ones that are flying in the same direction as one flock, and those that are flying in a second direction as a second flock. This is an instance of what the Gestalt psychologists called the principle of "common fate" – the perceptual system groups those sensory inputs that are changing in the same way at the same time.

A similar principle operates in hearing. An example occurs when a number of frequency components change in intensity at the same time, for example, when some become suddenly louder in a synchronized way. When this happens, these components are grouped together (perceptually "fused") and treated as parts of a single sound. Their combination may define a particular pitch or timbre. Two other similarities of frequency components that are present at the same time can also cause them to be fused and thereby define a single sound: a) The components have a common origin in space; b) They are all multiples of a common lower frequency (a fundamental frequency). This latter principle is useful because most sounds that have a pitch (e.g., a voice, or a violin) contain a fundamental frequency and many other frequencies that are multiples of that fundamental. There are also other similarities that affect this perceptual fusion, some of which are known and probably others that are as yet unknown.

Effects of Simultaneous Organization

When Auditory Scene Analysis (ASA) sorts out components of the incoming mixture and allocates them to different perceived sounds, this influences many aspects of what we hear, because only the frequency components assigned to the same sound by ASA will affect the experienced qualities of that sound. Examples are the pitch and timbre of the sound, both of which are based on the set of harmonics assigned to that sound.

Even the loudness of sounds can be affected by their perceptual organization. When two soft sounds occur at the same time, their energies are added up at the ear of the listener, giving the same energy as a single loud signal. So when our ear receives that loud signal, the auditory system has to form an interpretation of what we are listening to (i.e., is it two or more soft sources of sound or one loud one?). The perceptual process makes that decision using the cues for separating concurrent sounds, and this gives rise to the loudness experience(s).

Perceived location in space can also be affected by grouping. When we receive sound waves from different directions, the auditory system must decide, for example, whether it has heard two sounds at different locations or one sound and a reflection of that sound from a nearby surface. If the auditory system decides that the latter has occurred, the two sources of sound (original and reflection) are grouped together and heard as a single sound coming from the location of the first-arriving components (i.e., the original sound).

Why Do We Use the Cues We Do?

Space limitations do not permit me to mention all the principles of grouping that allow the auditory system to recover separate descriptions of the various individual sound sources in the listener's environment. But even if we knew them all and could list them, we would still have to ask, "Why these particular principles?" For an answer, we have to turn to regularities in the "ecology" of sound: What kinds of sound are there? How does sound reach our ears? Do sounds tend to overlap in time? And so on. No matter what causes a sound, there are certain relations that tend to

be present among its components. The auditory system's processes of organization take advantage of these regularities. Here is an example: All the frequency components of a single sound start at approximately the same time. Therefore, if the auditory system detects (within a complex mixture) a number of frequency components that all start ~~at the~~ together (± 25 ms), it should assign them all to the same sound. Here is another example: Many animal sounds, including the voiced parts of human speech, such as vowels, tend to consist only of harmonics – that is, frequency components that are all related to the same fundamental. Therefore, if the auditory system detects a number of components all related to the same fundamental, it should fuse them together as a single sound, and remove their contribution from the mixture.

Why Use Multiple Cues. Why Not the Strongest Ones Only?

We have found in our research that the auditory system "adds up" a number of cues at the same time in converging on the "best" grouping of acoustic components. Why should it not just use a single strong cue such as the different spatial origins of the components? The answer is this: If it did, there would be some circumstances in which we could not segregate sounds that were separate in the environment. For example, the normal spatial cues for segregation are missing when the signal comes from a single-loudspeaker radio, or around a corner, or when two sound sources are close together (e.g., a singer and a guitar) so their acoustic components all seem to come from the same location. Fortunately, we have many other cues that tell us the right way to allocate the sound energy.

To conclude this section on ASA, I should mention that researchers in areas outside of psychology have demonstrated a strong interest in the psychological research on this topic. Because the perception of speech by human listeners must also be done in complex backgrounds (the so-called "cocktail party problem"), ASA has been applied to the recognition of speech. It has been shown by speech scientists and auditory scientists that many ASA principles operate in this process. This has elicited a strong interest from engineers and computer scientists who are working on computer programs that can solve the ASA problem (Rosenthal & Okuno, 1998). Those researchers who think that their computer methods should incorporate the ones used by the human listener have called their field *"computational auditory scene analysis,"* or *CASA*. There is a strong practical reason for designing computer systems that can carry out ASA. Current attempts at having computers recognize human speech tend to fail when the talker is speaking in a background of other sounds. If a computer program could segregate the speech of a target talker from background sounds (including other talkers), the recognition process would not mix up irrelevant sounds with the relevant ones.

Neuroscientists have also become interested in ASA (e.g., Grossberg, Govendarajan, Wyse, & Cohen, 2004). Since nonhuman animals also face a world in which sounds are mixed at their ears, scientists have begun to study ASA in animals (e.g., Moss & Surlykke, 2001). Because they can intervene experimentally in the brain processes of animals, researchers hope to identify the brain processes that carry out ASA. Others have begun to study how the human brain does it, using various recording techniques such as *functional magnetic resonance imaging (fMRI)* and *EEG evoked potentials* (e.g., Izenberg & Alain, 2003).

There has also been an interest in ASA by music theorists and composers (e.g., Mountain, 1993). The principles discovered in the psychologist's laboratory seem to mirror many principles of musical composition, giving the latter a scientific foundation. Researchers in Hearing Science have also been studying ASA because even when fitted with a hearing aid or cochlear implant that produces a clearly audible signal, the user may still have great difficulty when more than one person is talking (the cocktail party problem). It is hoped that some modification of hearing aids or implants will allow their users to deal with this problem. The research on ASA has also attracted the attention of audio engineers, because their job is to control the blending of sounds in the recording or reinforcement of musical performances (Bregman & Woszczyk, 2004; Woszczyk & Bregman, in press).

Objective and Subjective Methods Revisited

In my laboratory, the use of our own ears – the philosopher's name is phenomenology – has greatly speeded up the process of discovering the ASA principles. Similarly, the use of demonstrations has strongly contributed to the acceptance of our theoretical claims. But what about other research in psychology? Most experimental psychologists claim that they work within an entirely objective methodology that is based on scientific principles. Historically, they have criticized clinicians for their use of intuition. In actual truth, the experiments done by *all* psychologists are built on a framework of intuition. An example is the fact that the researcher does not give a sly wink at the subjects when they sit down to participate in a psychology experiment (unless this is part of the protocol). How do the researchers know not to do this? Are they basing their actions on scientific principles (e.g., a

theory of what constitutes an act of communication, a theory about the interpretation of gestures, a lexicon of 1,000 different human gestures and the meaning of each in a variety of different contexts)? Not at all. Their actions are based on their personal experiences as human beings, from which they have derived an unformalized understanding about how people react to social signals, an understanding that is shared by all the members of their culture. A space alien would not know these things, so psychological research on humans would be a thousand times harder for it to do. Yet among human researchers, this priceless intuitive knowledge is concealed from the public eye, and in the final written report, objectivity is made to look like it is the only governing principle.

Address comments about this article to Dr. A. S. Bregman, Department of Psychology, McGill University, 1205 Dr. Penfield Avenue, Montréal, Québec, Canada H3A 1B1 (E-mail: al.bregman@mcgill.ca).

Résumé

Le présent article traite de deux sujets : a) le rôle de la subjectivité dans la recherche en psychologie, et b) mes travaux de recherche sur l'organisation perceptuelle du son dans lesquels la subjectivité a joué un rôle important. Des démonstrations sonores, attrayantes au regard de l'expérience subjective du lecteur, sont présentées au lieu des données de recherche objectives visant à appuyer les affirmations expliquant l'organisation auditive. Nous faisons valoir que tous les travaux de recherche en psychologie dépendent d'un cadre sous-jacent reposant sur l'intuition (des connaissances non formelles acquises grâce à l'expérience quotidienne) et que l'intuition joue un rôle dans la conception des expériences.

References

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound.* Cambridge, MA: Bradford Books, MIT Press. (Paperback 1994)

Bregman, A. S., & Ahad, P. (1996). *Demonstrations of auditory scene analysis: The perceptual organization of sound.* Audio compact disk. (Distributed by MIT Press).

Bregman, A. S., & Woszczyk, W. (2004) Controlling the perceptual organization of sound: Guidelines derived from principles of auditory scene analysis. In K. Greenebaum & R. Barzel (Eds.), *Audio anecdotes: Tools, tips and techniques for digital audio.* Natick, MA: A. K. Peters

Dannenbring, G. L. (1976). Perceived auditory continuity with alternately rising and falling frequency transitions. *Canadian Journal of Psychology, 30*, 99-114.

Grossberg, S., Govendarajan, K. K., Wyse, L. L., Cohen, M. A. (2004). ARTSTREAM: A neural network model of auditory scene analysis and source segregation. *Neural Networks, 17(4)*, 511-536.

Guzman, A. (1969). Decomposition of a visual scene into three-dimensional bodies. In A. Grasselli (Ed.), *Automatic interpretation and classification of images* (pp. 243-276). New York: Academic Press.

Izenberg, A., & Alain, C. (2003). Effects of attentional load on auditory scene analysis. *Journal of Cognitive Neuroscience, 15*, 1063-1073.

Moss, C. F., & Surlykke, A. (2001) Auditory scene analysis by echolocation in bats. *Journal of the Acoustical Society of America, 110*(4), 2207-26.

Mountain, R. (1993). *An investigation of periodicity in music, with reference to three twentieth-century compositions: Bartok's Music for Strings, Percussion, & Celesta; Lutoslawski's Concerto for Orchestra; Ligeti's Chamber Concerto.* PhD dissertation, School of Music, University of Victoria, BC.

Rosenthal, D. F., & Okuno, H. G. (Ed.). (1998). *Computational auditory scene analysis.* Mahwah, NJ: Lawrence Erlbaum Associates.

Winston, P. H., (Ed.). (1975). *The psychology of computer vision.* New York: McGraw-Hill.

Woszczyk, W., & Bregman, A. S. (in press). Creating mixtures: The application of auditory scene analysis (ASA) to audio recording. In K. Greenebaum & R. Barzel (Eds.), *Audio anecdotes.* (Vol 2). Natick, MA: A. K. Peters.